**Final Scientific Report**

**Cover Page**

**BARD Project Number:** IS-4394-11 R

**Date of Submission of the report: March 1, 2015**

**Project Title:** Optimization of methodology for genomic selection of moderate and large dairy cattle populations


| **Investigators** | **Institutions** |
|---|---|
| **Principal Investigator (PI):** Joel I. Weller | Agricultural Research Organization |
| **Co-Principal Investigator (Co-PI):** Ignacy Misztal | University of Georgia |
| **Collaborating Investigators:** Micha Ron | Agricultural Research Organization |


---


**Keywords** *not* appearing in the title and in order of importance. Avoid abbreviations.


**Abbreviations commonly** used in the report, in alphabetical order**:**
APGD = *a posteriori* granddaughter design, **EBV=**estimated breeding value, **GEBV=** genomic estimated breeding value, **SNP**= single nucleotide polymorphism, QTL= quantitative trait loci.

**Budget:** IS: $145,000 US: $145,000 Total: $290,000

_____ _____

Signature
Principal Investigator

Signature
Authorizing Official, Principal Institution

## Final Scientific Report

**Publication Summary** (numbers)

| | Joint IS/US authorship | US Authors Only | Israeli Authors Only | Total |
|---|---|---|---|---|
| Refereed (published, in press, accepted) BARD support acknowledged | 2 | 4 | 7 | 13 |
| Submitted, in review, in preparation | | | | |
| Invited review papers | | | 1 | 1 |
| Book chapters | | | 1 | 1 |
| Books | | | | |
| Master theses | | | | |
| Ph.D. theses | | | 1 | 1 |
| Abstracts | 2 | 10 | 4 | 16 |
| Not refereed (proceedings, reports, etc.) | 1 | 8 | 3 | 12 |

**Postdoctoral Training:** List the names and social security/identity numbers of all postdocs who received more than 50% of their funding by the grant.

Daniela Lino (does she have a SS#?)

**Cooperation Summary (numbers)**

| | From US to Israel | From Israel to US | Together, elsewhere | Total |
|---|---|---|---|---|
| Short Visits & Meetings | 1 | 1 | 1 | 3 |
| Longer Visits (Sabbaticals) | | 1 | | 1 |

**Patent Summary** (numbers)

| | Israeli inventor only | US inventor only | Joint IS/US inventors | Total |
|---|---|---|---|---|
| Submitted | | | | |
| Issued (allowed) | | | | |
| Licensed | | | | |

**Final Scientific Report**
**Abstract**

The main objectives of this research was to detect the specific polymorphisms responsible for observed quantitative trait loci and develop optimal strategies for genomic evaluations and selection for moderate (Israel) and large (US) dairy cattle populations. A joint evaluation using all phenotypic, pedigree, and genomic data is the optimal strategy. The specific objectives were: 1) to apply strategies for determination of the causative polymorphisms based on the "*a posteriori* granddaughter design" (APGD), 2) to develop methods to derive unbiased estimates of gene effects derived from SNP chips analyses, 3) to derive optimal single-stage methods to estimate breeding values of animals based on marker, phenotypic and pedigree data, 4) to extend these methods to multi-trait genetic evaluations and 5) to evaluate the results of long-term genomic selection, as compared to traditional selection. Nearly all of these objectives were met. The major achievements were:

1. The APGD and the modified granddaughter designs were applied to the US Holstein population, and regions harboring segregating quantitative trait loci (QTL) were identified for all economic traits of interest.
2. The APGD was able to find segregating QTL for all the economic traits analyzed, and confidence intervals for QTL location ranged from ~5 to 35 million base pairs.
3. Genomic estimated breeding values (GEBV) for milk production traits in the Israeli Holstein population were computed by the single-step method and compared to results for the two-step method.
4. The single-step method was extended to derive GEBV for multi-parity evaluation.
5. Long-term analysis of genomic selection demonstrated that inclusion of pedigree data from previous generations may result in less accurate GEBV.

Major conclusions are:

1. Predictions using single-step genomic best linear unbiased prediction (GBLUP) were the least biased, and that method appears to be the best tool for genomic evaluation of a small population, as it automatically accounts for parental index and allows for inclusion of female genomic information without additional steps.
2. None of the methods applied to the Israeli Holstein population were able to derive GEBV for young bulls that were significantly better than parent averages. Thus we confirm previous studies that the main limiting factor for the accuracy of GEBV is the number of bulls with genotypes and progeny tests.
3. Although 36 of the grandsires included in the APGD were genotyped for the BovineHD BeadChip, which includes 777,000 SNPs, we were not able to determine the causative polymorphism for any of the detected QTL.
4. The number of valid unique markers on the BovineHD BeadChip is not sufficient for a reasonable probability to find the causative polymorphisms. Complete resequencing of the genome of approximately 50 bulls will be required, but this could not be accomplished within the framework of the current project due to funding constraints.
5. Inclusion of pedigree data from older generations in the derivation of GEBV may result is less accurate evaluations.

**Genomic evaluation of the Israeli dairy population and effect of genotyped cow information in multi-parity analyses**

Methods for genomic prediction were evaluated for an Israeli Holstein dairy population of 829,437 cows and 1,305 progeny-tested bulls with genotypes. Inclusion of genotypes of 343 elite cows in an evaluation method that considers pedigree, phenotypes, and genotypes simultaneously also was evaluated. Two data sets were available: a complete data set with production records from 1985 through 2011, and a reduced data set with production after 2006 deleted. For each production trait, a multi-trait animal model was used to compute traditional genetic evaluation for parities 1 through 3 as separate traits. Evaluations were calculated for the reduced and complete data sets. Daughter deviations in 2011 were used as the expected future progeny performance of 135 validation bulls.

The estimated breeding values from the reduced data set were used to calculate parent average for validation bulls, which was the benchmark for comparing gain in predictive ability from genomics. Genomic predictions for bulls in 2006 were calculated using a Bayesian linear regression method (BayesC), genomic BLUP (GBLUP), single-step GBLUP (ssGBLUP), and weighted ssGBLUP (WssGBLUP). Predictions using BayesC and GBLUP were calculated either with or without an index that included parent average. Genomic predictions that included elite cow genotypes were calculated using ssGBLUP and WssGBLUP. Predictive ability was assessed by coefficients of determination ($R^2$) and regressions of predictions of validation bulls with no daughters in 2006 on daughter deviations of those bulls in 2011.

A reduction in correlation, $R^2$, and regression was observed through parities. Fat and protein yields had the lowest $R^2$ for all the methods. On average, $R^2$ was lowest for PA followed by GBLUP, BayesC, ssGBLUP, and WssGBLUP. For several traits, $R^2$ for direct genomic values from BayesC and GBLUP were lower than those for PA. Predictions using ssGBLUP were the least biased, and that method appears to be the best tool for genomic evaluation of a small population, as it automatically accounts for parental index and allows for inclusion of female genomic information without additional steps. Weighted ssGBLUP has the potential for higher evaluation accuracy. Accuracy of genomic evaluations is dependent on model refinement.

Past reports indicated sensitivity of EBV to definitions of unknown parents. Additional reranking of the groups could be expected by the use of single-step GBLUP. We obtained

solution to unknown parent groups by both BLUP and ssGBLUP. The groups formed 3 clusters corresponding to their definitions. Fluctuations over years among groups in BLUP were mostly small. Changes in solutions for groups in ssGBLUP were also small except for one group. The model for analysis of Israeli dairy data can be somewhat improved by modifying selected definitions of unknown parent groups. Genomic EBV were converted to estimates of SNP effects via the GWAS procedure of ssGBLUP. In this procedure genomic EBV were converted to SNP effects, SNP effects are used to calculate SNP variances, and the variances are then applied to a genomic relationship matrix. The procedure is iterative. We obtained Manhattan plots for 3 parities of 5 traits. The plots indicated a few large QTL. However the values of large SNPs vary considerably among parities. We are refining procedures for GWAS with ssGBLUP to validate the results.

**Application of a posteriori granddaughter and modified granddaughter designs to the US Holstein dairy population**

The APDG and modified granddaughter designs were applied to determine haplotype effects for Holstein bulls and cows with BovineSNP50 genotypes. The APDG was applied to 52 sire families, each with ≥100 genotyped sons with genetic evaluations based on progeny tests. For 33 traits (milk, fat, and protein yields; fat and protein percentages; somatic cell score; productive life; daughter pregnancy rate; heifer and cow conception rates; service-sire and daughter calving ease; service-sire and daughter stillbirth; 18 conformation traits; and net merit), the analysis was applied to the autosomal segment with the single nucleotide polymorphism (SNP) with the greatest effect in the genomic evaluation of each trait. All traits except two had a within-family haplotype effect. The same design was applied with the genetic evaluations of sons corrected for SNP effects associated with chromosomes besides the one under analysis. Number of within-family contrasts was 166 without adjustment and 211 with adjustment. Of the 52 bulls analyzed, 36 had BovineHD BeadChip genotypes that were used to test for concordance between sire quantitative trait loci and SNP genotypes. Complete concordance was not obtained for any effects.

Of the 31 traits with effects from the a posteriori granddaughter design, 21 were analyzed with the modified granddaughter design. Only sires with a contrast for the a posteriori granddaughter design and ≥200 granddaughters with a record usable for genetic evaluation were included. Calving traits could not be analyzed because individual cow evaluations were not computed. Eight traits had within-family haplotype effects. With respect to milk and fat

yields and fat percentage, the results on BTA 14 corresponded to the hypothesis that a missense mutation in *DGAT1* is the main causative mutation, although other polymorphisms in that gene also modify fat yield and percentage. The positive allele for protein concentration was less frequent, which indicated that selection on that locus could be effective. Although the results can be used to determine causative polymorphisms for most of the analyzed traits, complete DNA sequencing of most of the analyzed sires probably will be required.

Of 617 haplotype segments spanning the entire bovine genome and each including ~5 × $10^6$ bp, 5 cMorgans, and 50 genes, 608 autosomal segments were analyzed. The statistical model of Weller et al. (2013) was used for each haplotype segment. For all 33 traits, there was at least one chromosomal region in which the nominal probability for the haplotype effect was < $10^{-8}$, which corresponds to genome-wide significance of <$10^{-4}$ (Lander and Kruglyak, 1995). Net merit, the main US selection index, had 7 chromosomes with nominal probabilities of <$10^{-8}$. For each of those putative QTL, at least 1 grandsire family had a within-family contrast with a *t*-value of >3.

Confidence intervals estimated by the nonparametric bootstrap for the largest effect for each of 9 traits are presented in Table 1. The bootstrap distribution generated by 100 samples was bimodal only for net merit, which had the widest 90% confidence interval (8 haplotype segments). This may be due to the fact that net merit is a composite trait. For all other chromosomes, the confidence interval spanned less than a third of the chromosome. The narrowest confidence interval (a single haplotype segment) was found for somatic cell score.

**Table 1.** Confidence intervals (CI) of 90% and nominal *P* ($-\log_{10}$) for selected effects using a nonparametric bootstrap

| Trait | Chromosome | Haplotype segment[1] | *P* ($-\log_{10}$) | Haplotype segments on chromosome | CI |
|---|---|---|---|---|---|
| Milk | 15 | 389 | 9.1 | 377–396 | 387–392 |
| Protein | 10 | 264 | 10.8 | 264–288 | 264–265 |
| Somatic cell score | 6 | 177 | 25.0 | 156–185 | 177 |
| Net merit[2] | 18 | 441 | 17.4 | 438–450 | 440–447 |
| Productive life | 6 | 177 | 22.4 | 156–185 | 176–179 |
| Daughter pregnancy rate | 18 | 445 | 16.1 | 438–450 | 444–448 |
| Heifer conception rate | 6 | 176 | 13.1 | 156–185 | 177–183 |
| Cow conception rate | 6 | 177 | 31.6 | 156–185 | 177–179 |
| Final score | 5 | 149 | 15.6 | 131–155 | 147–151 |

[1]The genome was divided into 617 segments of ~75 markers each. Haplotype segments were numbered consecutively beginning with BTA1 and concluding with the sex chromosome.

[2]Bimodal distribution for bootstrap results.

*Details of cooperation*: We were in nearly constant E-mail communication. Dr. Weller spent a four month Sabbatical in the US from October 2012 through February 2013. Dr. Weller made a short visit to the US in October 2013. Dr. Misztal made a short visit to Israel in April 2014. We also met to finalize the research at the World Congress of Genetics Applied to Animal Production in Vancouver, in August, 2014. The Israeli data was analyzed by the US group, and two joint publications were published in the *Journal of Dairy Science*.

## Publications in Reviewed Journals

Glick, G. Shirak, A., Uliel, S. Zeron, Y., Ezra, E., Seroussi, E., Ron, M., and Weller, J. I. 2012. Signatures of contemporary selection in the Israeli Holstein dairy cattle. *Anim. Genet.* **43** (Suppl. 1), 45–55. (Invited paper).

Weller, J. I., Ron, M., Glick, G., Shirak, A. Zeron, Y., and Ezra, E. 2012b. A simple method for genomic selection for moderately sized dairy cattle populations. *Animal* 6, 193-202.

Misztal, I., Z.G. Vitezica, A. Legarra, I. Aguilar, and A.A. Swan. 2013. Unknown-parent groups in single-step genomic evaluation. *J. Anim. Breed. Genet*. **130**; 252–258.

Seroussi, E., Glick, G., Shirak, A., Ezra, E., Zeron, Y., Ron, M., and Weller, J. I. 2013. Maternity validation using sire-only BovineSNP50 BeadChip data. *Anim. Genet.* **44**; 754–757.

Weller, J. I., VanRaden, P. M., and Wiggans, G. R. 2013. Application of a posteriori granddaughter and modified granddaughter designs to determine Holstein haplotype effects. *J. Dairy Sci.* **96**; 5376–5387.

Lourenco, D.A.L., Misztal, I., Tsuruta, S., Aguilar, I., Ezra, E., Ron, M., Shirak, A., and Weller, J. I. 2014. Methods for genomic evaluation in a relatively small dairy population and the impact of inclusion of genotyped cows' information in multiple-parity analyses. *J. Dairy Sci.* **97**; 1742-1752**.**

Lourenco, D.A.L., Misztal, I., Tsuruta, S., Aguilar, Lawlor, T. J., Forni, S., and Weller, J. I. 2014. Are evaluations on young genotyped animals benefiting from the past generations? *J. Dairy Sci.* **97;** 3930-3942.

Misztal, I., A. Legarra, and I. Aguilar. 2014. Using recursion to compute the inverse of the genomic relationship matrix. *J. Dairy Sci.* **97**; 3943-3952.

Tsuruta, S., I. Misztal, D. A. L. Lourenco, and T. J. Lawlor. 2014. Assigning unknown parent groups to reduce bias in genomic evaluations of final score in US Holsteins. *J. Dairy Sci.* **97**; 5814-5821.

Wang, H., I. Misztal, I. Aguilar, A. Legarra, R. L. Fernando, Z. Vitezica, R. Okimoto, T. Wing, R. Hawken, and W. M. Muir. 2014. Genome-wide association mapping including phenotypes from relatives without genotypes in a single-step (ssGWAS) for 6-week body weight in broiler chickens. *Frontiers Genet*. DOI=10.3389/fgene.2014.00134.

Weller, J. I., Cole, J. B., VanRaden, P.M., and Wiggans, G. R. 2014. Application of the a posteriori granddaughter design to the Holstein genome. *Animal* **8**; 511–519.

Weller, J. I., Glick, G., Shirak, A., Ezra, E., Seroussi, E., Shemesh, M., Zeron, Y., and Ron, M. 2014. Predictive ability of selected subsets of single nucleotide polymorphisms (SNPs) in a moderately sized dairy cattle population. *Animal* **8**; 208–216.

Weller, J. I., and Ezra, E. 2015. Environmental and genetic factors affecting cow survival of Israeli Holsteins. *J. Dairy Sci.* **98**; 676–684.